Red riding on hood: Exploring how galaxy colour depends on environment

Pankaj C. Bhambhani,[1]★ Ivan. K. Baldry,[1]† Sarah Brough,[2] Alexander D. Hill,[1,6] M. A. Lara-Lopez,[3] J. Loveday,[4] B.W. Holwerda,[5]

[1]*Astrophysics Research Institute, Liverpool John Moores University, IC2, Liverpool Science Park, 146 Brownlow Hill, Liverpool L3 5RF, UK*
[2]*School of Physics, University of New South Wales, NSW 2052, Australia*
[3]*Departamento de Física de la Tierra y Astrofísica, Instituto de Física de Partículas y del Cosmos, IPARCOS. Universidad Complutense de Madrid (UCM), E-28040, Madrid, Spain*
[4]*Astronomy Centre, University of Sussex, Falmer, Brighton BN1 9QH, UK*
[5]*Department of Physics and Astronomy, 102 Natural Science Building, University of Louisville, Louisville KY 40292, USA*
[6]*Department of Physics, University of Liverpool, Liverpool L69 7ZE, UK*

Draft version July 2022

ABSTRACT

Galaxy populations are known to exhibit a strong colour bimodality, corresponding to blue star-forming and red quiescent subpopulations. The relative abundance of the two populations has been found to vary with stellar mass and environment. In this paper, we explore the effect of environment considering different types of measurements. We choose a sample of 49, 911 galaxies with $0.05 < z < 0.18$ from the Galaxy And Mass Assembly survey. We study the dependence of the fraction of red galaxies on different measures of the local environment as well as the large-scale "geometric" environment defined by density gradients in the surrounding cosmic web. We find that the red galaxy fraction varies with the environment at fixed stellar mass. The red fraction depends more strongly on local environmental measures than on large-scale geometric environment measures. By comparing the different environmental densities, we show that no density measurement fully explains the observed environmental red fraction variation, suggesting the different measures of environmental density contain different information. We test whether the local environmental measures, when combined together, can explain all the observed environmental red fraction variation. The geometric environment has a small residual effect, and this effect is larger for voids than any other type of geometric environment. This could provide a test of the physics applied to cosmological-scale galaxy evolution simulations as it combines large-scale effects with local environmental impact.

Key words: galaxies: evolution – galaxies: statistics – galaxies: fundamental parameters

1 INTRODUCTION

The field of galaxy formation and evolution tackles the task of explaining how a universe with a spatial uniformity of 1 in 10^5 - as evident from the Cosmic Microwave Background (CMB) radiation and other observations - evolved into the complex population of galaxies we see today. This population exhibits strong variations in a diverse set of properties including stellar mass, luminosity, size, morphology, colour, star formation rate (SFR), mean stellar age, etc. (e.g. Strauss 2002; Peacock 2003; Driver et al. 2009)

The framework currently favoured by cosmologists for explaining these properties is the ΛCDM model. Here cold dark matter (CDM) haloes form hierarchically from local overdensities, and the baryons fall into the gravitational potential wells of these halos and form galaxies (e.g. Navarro et al. 1997). However, there are many details not well understood, especially due to the complex interactions of baryons with other baryons and with dark matter.

A wide variety of models attempting to produce realistic galaxy populations exist. These may be broadly separated into two classes: semi-analytic models (SAMs) and hydrodynamical simulations. SAMs predict the formation and evolution of galaxies by combining the merger trees of dark matter-only N-body simulations with analytic prescriptions of baryonic physics (e.g. Bower et al. 2006; Croton et al. 2006; Somerville et al. 2008; Lagos et al. 2012; Bower et al. 2017). Hydrodynamical simulations, in contrast, evolve dark matter and baryons simultaneously and broadly self-consistently (Schaye et al. 2010; Dubois et al. 2014; Hopkins et al. 2014; Vogelsberger et al. 2014; Schaye et al. 2015; Crain et al. 2015; Springel et al. 2018; Davé et al. 2019). Baryonic processes taking place un-

★ E-mail: pcb.astro@gmail.com
† E-mail: I.Baldry@ljmu.ac.uk

ORIGINAL UNEDITED MANUSCRIPT

© 2023 The Author(s) Published by Oxford University Press on behalf of Royal Astronomical Society

der the resolution limit of the simulation are modelled via analytic 'subgrid' routines.

Both classes of model are naturally informed by observations of galaxy populations. Recent large scale multi-wavelength galaxy surveys provide a wealth of observations which enable an improved understanding of galaxy formation and evolution.

1.1 Galaxy Colour Bimodality

The advent of large scale galaxy surveys have helped discover a range of features in galaxy properties. One such interesting feature is the bimodality in galaxy colour. Strateva et al. (2001) first observed this in the observed-frame colour using data from the Sloan Digital Sky Survey (SDSS, Stoughton et al. 2002). Baldry et al. (2004a) analyzed the effect in the rest-frame colour when looking at the colour magnitude relations using SDSS data.

The colour-magnitude diagram (CMD) is an important diagnostic relation in astronomy, particularly in galaxy evolution. The absolute magnitude of a galaxy is its integrated starlight and is a proxy for stellar mass. The colour encodes information on the ages of the contributing stellar populations, i.e. the star formation history – barring extinction caused by dust. A CMD thus helps understand the evolution of star formation in the galaxy as a function of its stellar mass. Recent works (e.g. Taylor et al. 2015) have used stellar population synthesis (SPS, Taylor et al. 2011a) to estimate the stellar mass, and have used colour-mass diagrams instead of colour-magnitude diagrams.

Subsequent research into colour bimodality has validated the findings of Strateva et al. (2001) and Baldry et al. (2004b) to at least $z < 2$ (e.g. Balogh et al. 2004; Bell et al. 2004; Williams et al. 2009; Taylor et al. 2015; Foltz et al. 2018). The colour bimodality is interesting to theorists as well because colour has a direct relation with star formation history.

When looking at just the colour of galaxies, the corresponding distribution appears to be well approximated as a sum of two Gaussian distributions (e.g. Strateva et al. 2001; Taylor et al. 2015). This suggests that galaxy formation processes given rise to two different dominant galaxy populations, each with a different distribution of colour values. The two populations are generally labeled "blue" and "red" based on their lower and higher colour values respectively, though their exact definitions have varied from one research work to another (e.g. Taylor et al. 2015). Studies of this bimodality at different redshifts have revealed another interesting finding - although the populations are roughly equivalent in total stellar mass at $z \sim 1$, the red population has nearly doubled in stellar mass, stellar mass density, and number density over the past ~ 7 Gyr (e.g Bell et al. 2004; Arnouts et al. 2007; Foltz et al. 2018). This suggests that galaxies move between the two populations - these are two evolutionary stages. But the existence of two distinct colour modes must mean the transition from blue to red stage is relatively quick (e.g. Balogh et al. 2004).

In order to reproduce the observed colour distribution and its evolution with redshift, theoretical models tend to add mechanisms that rapidly disrupt or prevent star formation in galaxies. This process is generically called "quenching" (see e.g. Peng et al. 2010), and it mainly involves removal of star forming gases. While the exact nature of quenching remains debated, popular candidates involve feedback from an Active Galactic Nuclei (AGN) (e.g. Benson et al. 2003; Bower et al. 2006; Croton et al. 2006; Somerville et al. 2008).

1.2 The Effect of Environment

In addition to stellar mass, galaxy colour is also found to be dependent on the local environment. Previous works have identified that galaxies are likely to be red in denser environments (e.g. Kauffmann et al. 2004a; Baldry et al. 2006; van der Burg et al. 2018; Reeves et al. 2021). Previous studies have found an increase in the local environment density to correlate with a decrease in the star-formation rate (Schaefer et al. 2017, 2019), a decrease in the number of star-forming galaxies (Barsanti et al. 2018) or a change in the stellar kinematic properties of galaxies (van de Sande et al. 2021). The most widely agreed reason for environmental dependence is therefore the loss of star forming gas as a result of stripping in denser environments (e.g. Barsanti et al. 2018; Trussler et al. 2020; Sotillo-Ramos et al. 2021). The major processes involved include galaxy harassment and mergers(e.g. Bialas et al. 2015), strangulation (e.g. Peng et al. 2015 and ram pressure stripping(Gunn & Gott 1972; Brough et al. 2013).

The environment of a galaxy, broadly speaking, is the region surrounding it that has a potential to interact with the galaxy. The term environmental density (also called an environmental measure) refers to a metric that characterises the kind of environment a galaxy is located in. Just like for the terms red and blue (as mentioned above), there are many commonly used definitions for environment. These include the number of (other) galaxies in a given region (e.g. Baldry et al. 2006; Yoon et al. 2008; Peng et al. 2010; Brough et al. 2013; Liske et al. 2015; Schaefer et al. 2017; Baldry et al. 2018; van de Sande et al. 2021), the type of large scale structure surrounding it (e.g. a void, a knot, etc, see Alpaslan et al. 2014; Eardley et al. 2015), group/cluster virial mass and distance from the cluster centre (e.g. Barsanti et al. 2018; van der Burg et al. 2018), position as a central/satellite galaxy in a halo (e.g. Balogh et al. 2016), and identifying galaxy groups using friends-of-friends based grouping algorithm (Robotham et al. 2011; Schaefer et al. 2019). The source of environmental dependence is generally thought to be the local overdensity (e.g. Baldry et al. 2006; Behroozi et al. 2019) but there may also be a possible effect of the region of the cosmic web in which the galaxy lies (e.g. Eardley et al. 2015). The former type of dependence is termed the local environment, while the latter is often termed the large-scale environment. In accordance with Eardley et al. (2015), we use the term use the term 'geometric environment' to denote the different regions of the cosmic web.

Baldry et al. (2006) investigated the effects of local environment on galaxy colour in detail. In particular, they studied how the double Gaussian fits to the colour distribution changed with the projected density of neighbouring galaxies. They found that increasing the environmental density has two effects - a major effect is the transition of galaxies from the blue Gaussian distribution to the red one, and a minor effect is the modest reddening of each Gaussian fit, i.e. the mean shifting to slightly higher colour values. They also found that within each environmental density bin, the data were always well modeled by a double Gaussian fit. This suggests that the galaxy properties within the blue and red distributions do not change much with environment.

In contrast, Alpaslan et al. (2015) investigated the effect of galaxy colour (and other properties) as a function of non-local environment. One of the environmental measures used by them was derived from classifications of the large scale structure, as obtained from the GAMA Large Scale Structure Catalogue (Alpaslan et al. 2014). Their results do not show a significant difference in galaxy colour (or any other property) when comparing across the different large scale structure environments.

ORIGINAL UNEDITED MANUSCRIPT

There has also been work done that compares the effect of different environments. Perez et al. (2009) study the effects of galaxy interactions and mergers in different local density and host-halo mass environments, using data from SDSS DR4. Wilman et al. (2005) obtain independent measurements of density on different scales and use that to study galaxy colour dependence in SDSS DR5 data. Pandey & Sarkar (2020) use spectroscopic data from SDSS DR16 to explore galaxy colour in different environments of the cosmic web.

Environmental effects in hydrodynamical simulations arise naturally from the evolution in the properties of individual gas particles. The change in these properties are computed through a combination of hydrodynamical schemes such as SPH (Hopkins 2013), and subgrid processes, e.g. star formation and radiative cooling (Wiersma et al. 2009). The influence of environmental processes such as ram-pressure stripping in quenching star-formation within simulated galaxies is a popular area of study (e.g. Trayford et al. 2016; Simpson et al. 2018). Discrepancy between simulations in the influence of environment arise due to differences in resolution, subgrid treatment, and the hydrodynamic solvers employed. SAMs, in contrast, require the inclusion of explicit analytic forms of environmental effects, such as galaxy mergers, tidal interactions and ram pressure stripping (Tecce et al. 2010; Ayromlou et al. 2019). The processes included and the methods employed vary between SAMs. The success of hydrodynamic simulations and SAMs are largely measured through their ability to match observations. With the advent of large-scale surveys like the SDSS and the Galaxy and Mass Assembly (GAMA) survey (Driver et al. 2009, 2011; Liske et al. 2015; Baldry et al. 2018), we have opportunities to study the environmental dependence in more detail and thereby further test the accuracy of simulations of galaxy formation and evolution.

There is also the potential to study the effect of geometric environment. The cosmic web (Bond et al. 1996) has been theorised to affect the properties of galaxies and their host dark matter haloes, with N-body simulations and simulations of galaxy formation and evolution playing a key role in this field of study (e.g. Hahn et al. 2007; Wang et al. 2011; Goh et al. 2019; Kraljic et al. 2020; Hellwing et al. 2021b). Simulations have also been used to explore the relationship between environment and galaxy properties (e.g. Bhowmick et al. 2020; Kristensen et al. 2021), as well as the variation in this relationship imparted by different measures of environment (Haas et al. 2012; Muldrew et al. 2012). It is suggested that measures of environment defined on varied spatial scales are encoded with different information, affected as they are by physical processes to a greater or lesser extent, which then leads to varied correlations with galaxy properties.

1.3 Aim of this work

In this paper, we examine the effect of environment and stellar mass on the galaxy colour distribution. We use as a metric the fraction of galaxies within a given mass/environment region that are red. We aim to answer the following questions:

- How does the red fraction change as a function of stellar mass and environment?
- Do the different environmental measures convey different information?
- In particular, can the local environment fully explain the red fraction variation, or is there a residual effect from the geometric environment (i.e. the cosmic web)?

In this sense, our high-level goals are similar to that of Baldry et al. (2006); Peng et al. (2010); Alpaslan et al. (2015) and Taylor et al. (2015). We build upon their work, by using the latest data and experimenting with different density measures. The remainder of this paper is organised as follows: section 2 describes the samples of galaxies used in the investigation, which is taken from the GAMA survey. Section 3 details the methodology used to compute the red fraction estimates. In section 4, we present and discuss our findings. Section 5 provides a summary of this work and suggests future research directions.

Throughout this work, we assume a flat ΛCDM cosmology with the parameters $H_0 = 70$ km/s/Mpc, $\Omega_{m,0} = 0.3$, $\Omega_{\Lambda,0} = 0.699$. All stellar mass estimates obtained from GAMA assume a Chabrier (2003) initial mass function (IMF). All magnitudes provided by GAMA are given in the AB magnitude system. Logarithms are assumed to be with base 10 unless specified otherwise, and abbreviated as log.

2 DATA

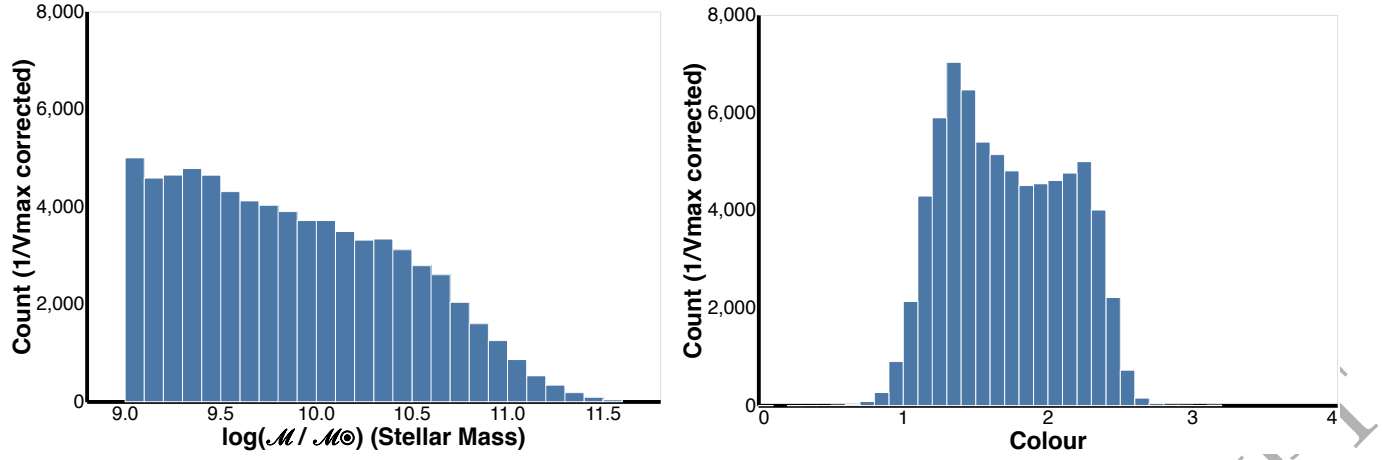
The Galaxy and Mass Assembly (GAMA) survey (Driver et al. 2009, 2011; Liske et al. 2015; Baldry et al. 2018; Driver et al. 2022) is an international large-scale galaxy redshift survey. Its goal is to probe the universe on intermediate scales between 1 kpc and 1 Mpc. While cosmological surveys have improved our understanding of the universe on large scales (> 1 Mpc), the picture is less clear on the scales of galaxies, groups and clusters. The goal of GAMA is to provide observations of structure formation at these scales, including measuring the dark matter halo mass function (HMF), as well as the effect of baryonic feedback through measurements of stellar mass function (SMF) down to very low mass haloes.

GAMA surveys the sky over 5 regions (three equatorial – named G09, G12 and G15 and two southern – named G02 and G23), covering a total area of 286 deg^2. It provides spectra (Hopkins et al. 2013) and redshift (Baldry et al. 2014) estimates from the AAOmega spectrograph as part of the 3.9m Anglo Australian Telescope. It also provides photometric data comprising of optical *ugriz* imaging from SDSS (DR7) and near-infrared (NIR) *ZYJHK* imaging from the VIKING survey. When analyzing environmental effects, it is important to ensure spectroscopic completeness in the sample to avoid selection effects due to local density of galaxies. Through multiple visits to each field (Robotham et al. 2010), GAMA provides accurate redshift estimates (> 98% confidence) with uniform and near total spectroscopic completeness (≳ 98%) upto a Petrosian r-band magnitude limit of $r_{petro} < 19.8$ mag, except in G23 which instead has an i-band limit of 19.2 mag. (Liske et al. 2015; Baldry et al. 2018).

2.1 Our Chosen Sample

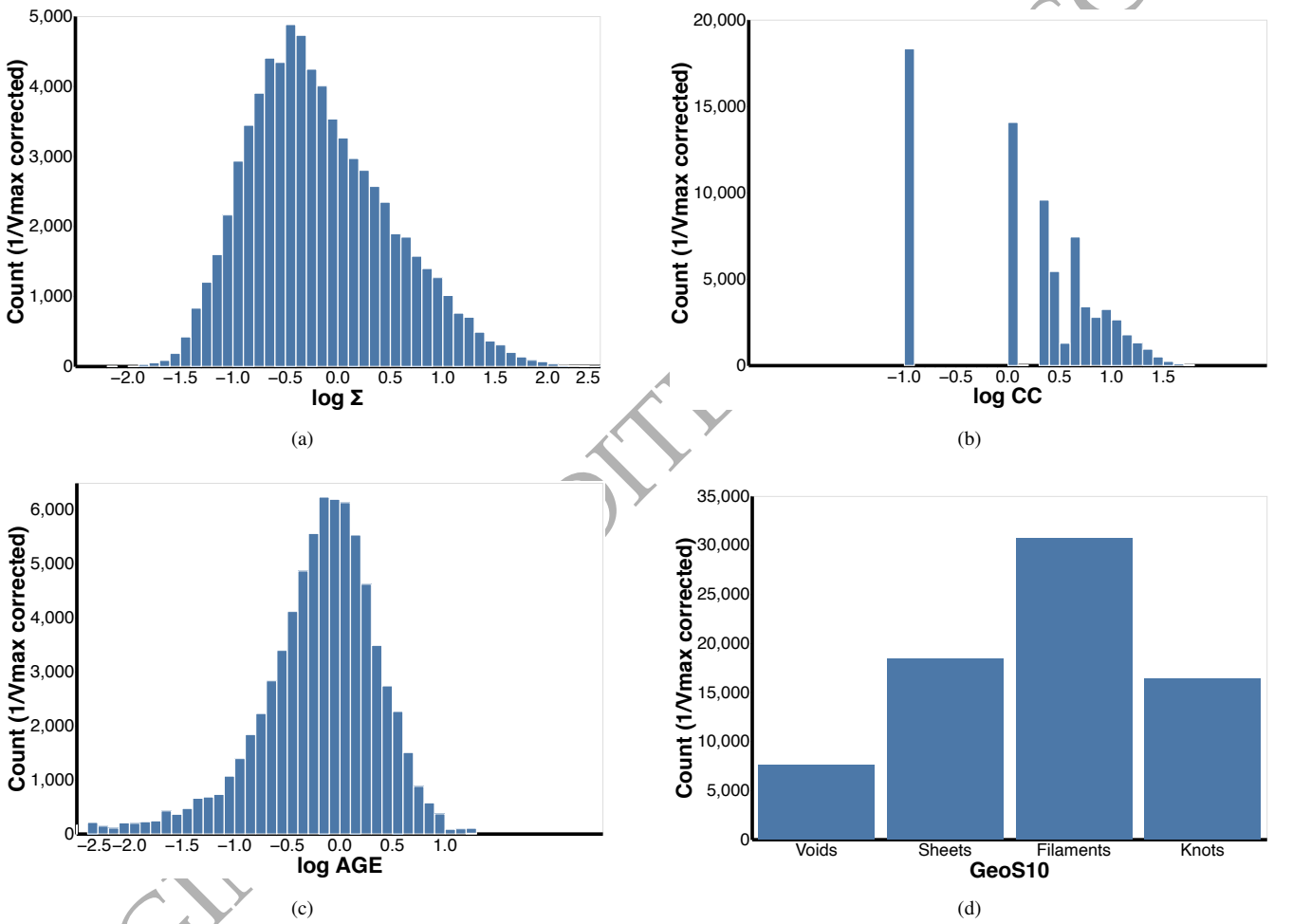
Our sample is obtained from Data Release 4 (DR4, Driver et al. 2022) of the GAMA survey. DR4 has redshift and other information for ∼ 330,000 objects. For the local environment, we choose the environment measures defined in the `EnvironmentMeasures` Data Management Unit (DMU) (Brough et al. 2013; Liske et al. 2015). Galaxy environment has been studied previously for GAMA using data products available in the GAMA Galaxy Group Catalogue (G3C, data provided in the `GroupFinding` DMU), where the neighbours of a galaxies are determined by a Friends of Friends (FoF) linking algorithm. (Robotham et al. 2011). In a FoF approach, all objects grouped together will have the same number of neigh-

ORIGINAL UNEDITED MANUSCRIPT



(a) (b)

Figure 1. Histograms of (a) mass and (b) colour attributes for the objects in our chosen sample. The counts shown here are volume ($1/V_{max}$) weighted, see section 3.1 for more information on how this correction is done.



(a) (b) (c) (d)

Figure 2. Histograms of the different environmental density measures for the objects in our chosen sample, namely (a) 5th Nearest Neighbour (Σ), (b) Cylindrical Count (CC), (c) Adaptive Gaussian Ellipsoid (AGE) and (d) Geometric Environment with count smoothing scale $\sigma = 10h^{-1}$ Mpc (GeoS10). See section 2.5 for details on these environmental measures. The counts shown here are volume ($1/V_{max}$) weighted, see section 3.1 for more information on how this correction is done.

bours. Since we are more interested in environmental measures that vary continuously, we do not include data from G3C in our sample.

We chose limits of $10^9\mathcal{M}_\odot \lesssim \mathcal{M} \lesssim 10^{11.6}\mathcal{M}_\odot$ and $0.05 \lesssim z \lesssim 0.18$, where $\mathcal{M}$ and z are the stellar mass and CMB-frame redshift respectively. See sections 2.2 and 2.4 for justification of our choice of limits for mass and redshift respectively. Our sample size is 49, 911 objects, about 16.6% of all objects in GAMA.

Below we provide a brief description of the different attributes in our samples.

2.2 Stellar Mass ($\mathcal{M}/\mathcal{M}_\odot$)

This is the total stellar mass of a galaxy in units of solar mass. It is estimated from stellar population synthesis (SPS) models of the measured galaxy spectral energy distributions (SEDs). For more details, see Taylor et al. (2011a) (and section 5.3.2 of Baldry et al. 2018). For our sample, we chose a lower limit of $10^9\mathcal{M}_\odot$ in line with Baldry et al. (2006), and an upper limit of $10^{11.6}\mathcal{M}_\odot$ due to lack of significant number of higher mass objects. Figure 1a shows the stellar mass histogram for objects in our sample.

2.3 Colour (u - r)

The galaxy colour used here is the difference between the u and r band magnitudes and is denoted $u-r$. This colour is obtained from SED fitting (Taylor et al. 2011b), with apertures obtained using LAMBDAR (Lambda Adaptive Multi-Band Deblending Algorithm in R) photometry (Wright et al. 2016). The values are provided as part of GAMA DR4 and are k-corrected as well as Milky Way dust extinction corrected. Figure 1b shows a histogram of the resulting colour values.

We also experimented with two other versions of colour, derived from the same bands but with variations in photometry and dust extinction correction. We defer to Appendix A the discussion of how the different colours affect the red galaxy fraction evolution. We choose the colour $u-r$ for work as it is derived from more accurate photometry compared to the other two colours and it produces a red fraction dependence consistent with what is known in the literature.

2.4 Redshift

There are two redshift attributes in our data set. One is the redshift z of the object, corrected to the CMB frame. We select $0.05 \lesssim z \lesssim 0.18$ for our chosen sample. The lower limit is to avoid regions dominated by individual stars. The upper limit is the maximum redshift for which the local environmental measurements are available in GAMA DR4. See section 2.5 for how this redshift upper limit is determined as well for definitions of environmental densities.

The other attribute is the limiting redshift z_{max} - the maximum redshift at which the object would be visible for a given r-band (Petrosian) magnitude limit. GAMA provides these estimates from iteratively solving the k-correction to a limiting magnitude of $r < 19.8$ mag. The limiting redshift is helpful for doing volumetric corrections (see section 3.1).

2.5 Environmental Densities

We use four different environmental densities in our dataset as provided by GAMA and as described below. The first three estimate the number of galaxies within the local neighbourhood, while the geometric environmental measure is related to the location of the galaxy within the cosmic web. As described in section 1.2, the first three are measures of local environment, while the final is a measure of the large-scale environment.

We note that all the local environmental measures are defined on a density defining population (DDP) of galaxies which is also volume limited (see section 3.1). This population is defined as all galaxies with $M_r(z_{ref} = 0, Q = 0.78) < -20\,\mathrm{mag}$, where Q defines the expected evolution of M_r as a function of redshift (Loveday et al. 2012). Given GAMA's limiting magnitude of $r < 19.8\,\mathrm{mag}$, this leads to a redshift limit of $z \lesssim 0.18$ for these density mesures. See Brough et al. (2013) for more information.

We note that objects in our sample have estimates defined for each environmental measure. Also, we exclude any objects that do not have reliable density estimates due to survey edge effects.

2.5.1 *5th Nearest Neighbour Surface Density (Σ)*

This environmental measure estimates the projected 2D surface density based on the distance to the N-th nearest spectroscopically confirmed bright neighbouring galaxy (within the DDP) within a given redshift and absolute magnitude range. If this distance is d_N, then the surface density is given by

$$\Sigma = \frac{N}{\pi d_N^2} \tag{1}$$

GAMA uses $N = 5$, a redshift range $\pm\triangle zc = 1000$km/s and absolute magnitude $M_r < -20\,\mathrm{mag}$ (Brough et al. 2013). When one or more of the Nneighbours are beyond the survey edge, additional considerations are required and upper limits are estimated. However, we do not include any such galaxies in our dataset.

Figure 2a shows the distribution of Σ for the galaxies in our sample. Note that in this paper we always plot environment measures in logarithmic units, since we are interested in their effects on an order of magnitude scale.

2.5.2 *Cylindrical Count (CC)*

Similar to the 5th nearest neighbour surface density, this measure provides the number of (other) galaxies (within the DDP) within a cylindrical volume centered on the galaxy in question. It is given by

$$CC = \frac{N_{cyl}}{\bar{n}_{ref} V_{cyl}} \propto N_{cyl} \tag{2}$$

where N_{cyl} is the number of galaxies in the cylinder and V_{cyl} is the volume of the cylinder. The (co-moving) radius of the cylinder is taken as 1 Mpc, and the height is determined by the redshift range of ±1000km/s as before. $\bar{n}_{ref} = 0.00911 Mpc^{-3}$ is the average number density of the density defining population. See Liske et al. (2015) for more information. Since V_{cyl} and $\bar{n}_{ref}$ are the same for every object, we can use the galaxy count N_{cyl} as a proxy for the overdensity.

Figure 2b shows a histogram of the cylindrical counts. About 25% of the objects do not have any galaxies in their cylinder and so have $N_{cyl} = 0$ In order to include these zero counts when working in log units, we set $\log(CC) = -1$ when $CC = 0$, which is why the histogram has a huge bar centered on -1.

ORIGINAL UNEDITED MANUSCRIPT

2.5.3 Adaptive Gaussian Environment Density (AGE)

This density measure is equivalent to a weighted local volume density of galaxies, where closer galaxies receive more weight than more distant ones (Yoon et al. 2008). First, the neighbouring galaxies are identified (from the DDP) as those lying within an adaptive Gaussian ellipsoid defined by

$$\left(\frac{r_a}{3\sigma}\right)^2 + \left(\frac{r_z}{\mathsf{AGEScale} * 3\sigma}\right)^2 \leq 1 \quad (3)$$

where r_a, r_z are the distances from the centre of the ellipsoid (i.e. from the position of the galaxy in question) in the plane of sky and along the line-of-sight in co-moving Mpc, respectively, and σ = 2Mpc. AGEScale is the adaptive scaling factor, defined as

$$\mathsf{AGEScale} = 1 + (0.2\,n) \quad (4)$$

where n is the number of galaxies from the DDP within 2 Mpc. AGEScale is used to scale the value of sigma along the redshift axis by up to a factor of 3 (corresponding to n = 10) for the highest density environments to compensate for the "finger-of-God" effect (Schawinski et al. 2007; Thomas et al. 2010). The adaptive Gaussian density parameter is then computed as:

$$AGE = \frac{1}{\sqrt{2\pi}\sigma} \sum_i \exp\left[-\frac{1}{2}\left(\left(\frac{r_{a,i}}{\sigma}\right)^2 + \left(\frac{r_{z,i}}{\mathsf{AGEScale} * \sigma}\right)^2\right)\right] \quad (5)$$

Figure 2c shows a histogram of the Adaptive Gaussian Environment density.

2.5.4 Geometric Environment: GeoS10

This environmental measure identifies the cosmic web of large scale structure within the GAMA equatorial survey regions by classifying the geometric environment of each point in space as either a void, a sheet, a filament or a knot (Eardley et al. 2015). The label is assigned based on the number of dimensions along which the underlying matter is collapsing. Galaxies are used as a proxy to estimate the overall matter distribution. A Hessian (matrix of second-order partial derivatives) of the gravitational potential is computed, and eigenvalues of the Hessian give the directions along which the Hessian is changing. This matrix is also called the tidal tensor. The number of positive eigenvalues is the number of dimensions along which matter is collapsing. In practice, the only directions considered are ones where the collapse is happening reasonably quickly. This is equivalent to counting the number of eigenvalues higher than a certain threshold. The classification assigned is as follows :

Voids: all eigenvalues below the threshold ($\lambda_1 < \lambda_{th}$)
Sheets: one eigenvalue above the threshold ($\lambda_1 > \lambda_{th}, \lambda_2 < \lambda_{th}$).
Filaments: two eigenvalues above threshold ($\lambda_2 > \lambda_{th}, \lambda_3 < \lambda_{th}$).
Knots: all eigenvalues above the threshold ($\lambda_3 > \lambda_{th}$).

where $\lambda_1, \lambda_2, \lambda_3$ are the three eigenvalues, with $\lambda_3 < \lambda_2 < \lambda_1$. λ_{th} is the threshold mentioned above. Thus, a 'void' region means that there is no significant collapse happening along any direction – it is a region surrounded by similar or higher density regions in all directions. Similarly, a 'sheet' region means that matter is collapsing mainly along a single direction, and so on.

The original galaxy counts used to estimate the matter distribution are smoothed to minimize noise and remove non-linearities. The smoothing scale σ is also a free parameter in addition to the eigenvalue threshold λ_{th}. Eardley et al. (2015) chose values of σ and λ_{th} so as to divide the galaxies into the four regions (determined by the four labels above) as equally as possible, so that any measurements made from these maybe statistically significant. Two combinations of these parameters are chosen: $(\sigma, \lambda_{th}) = (4h^{-1}\mathsf{Mpc}, 0.4)$ or $(10h^{-1}\mathsf{Mpc}, 0.1)$. Both combinations have at least 10% of galaxies within each of the four regions. For our work, we want to use this measure to study environmental effects on large scales and be independent of the local environmental measures. Hence we choose the estimates from the larger smoothing scale of $10\,h^{-1}$ Mpc. We refer to it as GeoS10 in accordance with Eardley et al. (2015). Figure 2d shows the histogram of this environmental measure. Overall there is a statistically significant number of galaxies in each of the four bins, with filaments have the highest ($\sim 30,000$) and voids having the lowest (~ 7500) number of galaxies.

The definition of geometric galaxy environment employed in this work partitions space into regions defined as being voids, sheets, filaments or knots. Galaxies are labeled according to the region they reside in, e.g. a galaxy within a knot region is defined as a knot galaxy (see Eardley et al. 2015, their Figs. 2 and 3). The geometric environment is therefore discretely defined. Other methodologies, such as the DisPerSE algorithm (Sousbie 2011), directly construct cosmic webs such that they are skeleton structures that exist within space, rather than partitioning it directly. Continuous studies of environment, such as how galaxy properties vary with distance to the closest knot (e.g. Kraljic et al. 2018; Malavasi et al. 2022), are then possible. An extension of the Eardley et al. (2015) methodology to enable such studies here are beyond the scope of this work.

3 METHODOLOGY

Figure 3a shows a colour mass diagram (contour plot + heatmap) of galaxy counts. We can identify two peaks (shown by blue and red crosses in the figure), which would typically be associated with the separation of the galaxy population into a blue and red sequence (Tully et al. 1982; Strateva et al. 2001; Baldry et al. 2004a,b; Taylor et al. 2015). To understand the effect of mass and environment on galaxy colour, we first need to separate the red and blue galaxy populations (section 3.2), so that we can estimate the red galaxy fraction and study its evolution. Before we do that, we need make sure that the sample is volumetrically unbiased (section 3.1).

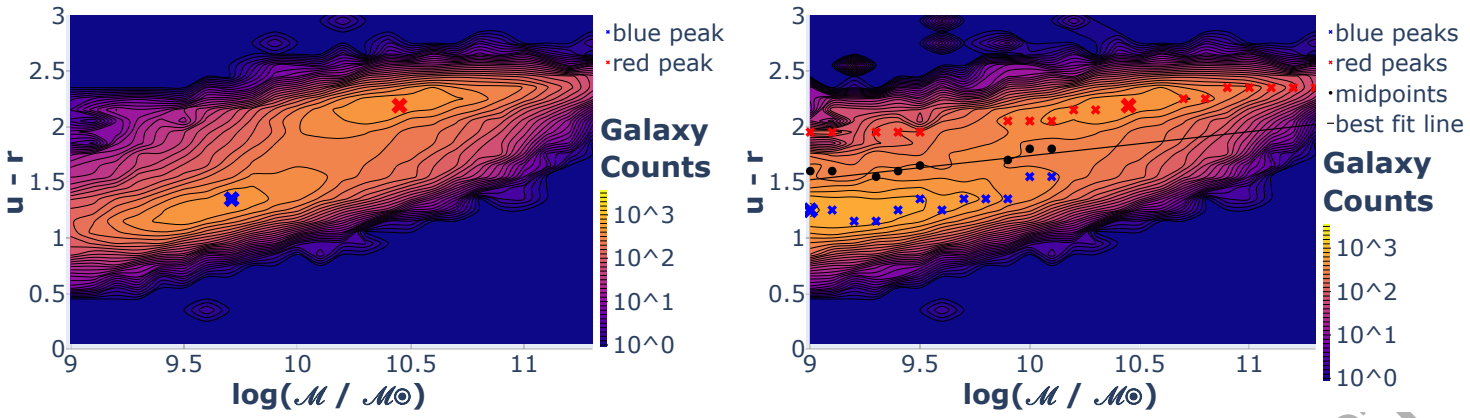
3.1 Volumetric Correction

3.1.1 The $1/V_{max}$ approach

Volumetric correction is important for our analysis because blue galaxies are brighter at a given stellar mass, and would therefore be overrepresented as compared to the red galaxies. This would lead to biased red fraction estimates.

We follow the approach of Felten (1976) to perform volumetric correction. Intuitively, as we observe out to higher redshifts, we will necessarily underestimate the number of faint galaxies, since the light reaching us from them is not sufficient for inclusion in the survey. The magnitude of a galaxy determines the redshift range over which it can be observed, and there may be a number of galaxies whose observable redshift range is much smaller than the survey redshift range. To correct for this, we weigh each galaxy by a factor $f = \frac{V_{survey}}{V_{max}}$ to make the sample complete (unbiased) within the volume, i.e. the chosen sky area and redshift limits. The term V_{max} is the maximum comoving volume over which the galaxy under

ORIGINAL UNEDITED MANUSCRIPT



(a) Without volumetric weighting

(b) volume weighted with dividing line

Figure 3. Contour plots and heatmaps representing the distribution of galaxy counts as function of the stellar mass $\log(\mathcal{M}/\mathcal{M}_\odot)$ and colour $u-r$. Panel (a) shows the raw counts, we see two peaks corresponding to red and blue region, shown as cross marks of the corresponding colour. Panel (b) shows volume $(1/V_{max})$ weighted counts. The corresponding peaks from Panel (a) are shown by large cross symbols. It also shows peaks in the red and blue region for each mass bin (represented as dots of the corresponding colour) along with midpoints (black dots) and the best line line of separation (black line).

concern can be observed within the redshift limits of the sample. For a galaxy, i, we compute $V_{max,i}$ as

$$V_{max,i} \propto [D(z_{max,i})^3 - D(z_{min,i})^3] \tag{6}$$

$D(z)$ is the comoving distance to redshift z . $z_{max,i}$ is the limiting redshift of the galaxy i (see section 2.4) at or the upper limit of the sample, whichever is lower. $z_{min,i} = z_{start}$ is the lower redshift limit of the sample.

The survey volume is computed in a similar manner

$$V_{survey} \propto [D(z_{limit})^3 - D(z_{start})^3] \tag{7}$$

where z_{limit} and z_{start} are the upper and lower redshift limits of the sample.

To estimate the limiting redshift (z_{max}), we use the distance modulus equation. Ignoring the evolution of the galaxy over the redshift range, we can write

$$\begin{aligned} m - \mu(z) - K(z) &= m_{limit} - \mu(z_{max}) - K(z_{max}) \\ \mu(z_{max}) &= \mu(z) + [m_{limit} - m] + [K(z) - K(z_{max})] \end{aligned} \tag{8}$$

which can be used to determine z_{max}. $\mu = 5\log_{10}(D_L/10pc)$ is the distance modulus (D_L is the luminosity distance), $K(z)$ is the K-correction. The differential K-correction $K(z_{max}) - K(z)$ is usually small and can be determined from the galaxy SED.

For our chosen sample, we have $z_{start} = 0.05$, $z_{limit} = 0.18$. We use the values of maximum redshifts z_{max} as provided by GAMA for a limiting Petrosian r-band magnitude of 19.8 mag, i.e. $m_{limit} = 19.8$ mag. These values are available under the column name `zmax_19p8` in the `STELLARMASSES` DMU (Driver et al. 2022).

The $1/V_{max}$ is a straightforward approach for working with data that is not volumetrically complete. It does however implicitly assume that the distribution of galaxies varies uniformly with redshift. Modifications to the method have been proposed to deal with the presence of large-scale structure, e.g. by introducing additional terms in the weighting factor V_{survey}/V_{max} (see Baldry et al. 2012).

3.1.2 *Clipping outliers*

Photometric errors can sometimes result in incorrect estimates of the limiting redshift, causing outliers to be present in the z_{max} distribution at a fixed mass. The outliers at lower values of z_{max} can cause more problems, since they would result in a higher values of $1/V_{max}$, thereby skewing up the overall galaxy counts. To avoid these outliers, we clip the z_{max} distribution to the 95th percentile within each (log) mass bin of width 0.01. The percentiles are computed for bins spaced at intervals of 0.25 and a second order polynomial curve is fitted to these points:

$$f(x) = a_0 + a_1 x + a_2 x^2 \tag{9}$$

The best fit values obtained are $(a_0, a_1, a_2) = (5.03, -1.1, 0.06)$.

The $1/V_{max}$ weights have values ranging from 1 – which is the median and the most common weight, to 18.6 – which occurs for low-mass galaxies ($\mathcal{M} \approx 10^9 \mathcal{M}_\odot$) The mean weight and standard deviation are both about 1.6.

Figure 3b shows the updated galaxy counts as a function of colour and stellar mass, corrected for volumetric completeness. We can see that while the red peak has stage at its original place, the blue peak has move towards lower mass $\mathcal{M} \approx 9\mathcal{M}_\odot$. Also, the correlation between mass and colour has reduced for the region around the blue peak, as compared to Figure 3a.

3.2 Separating the two populations

We can now use the $1/V_{max}$ weighted galaxy counts to separate the red and blue populations. We assume that the colour evolution of the population over our redshift range of interest is not significant enough to matter for the separation. The simplest approach for this would be to fit a line through the contour distribution (equivalently a plane through the 3D distribution) that divides the two populations well. For our purposes, it is not necessary for this line to be exact - an approximate line would add some noise when computing the red fraction within a given mass/environment bin, but it would still retain the overall signal of the variation in red fraction varies as a function of stellar mass and environmental density.

ORIGINAL UNEDITED MANUSCRIPT

Environmental Density	Red Fraction range (averaged across stellar mass)
5th Nearest Neighbour (Σ)	0.492 ± 0.007
Cylindrical Count (CC)	0.499 ± 0.007
Adaptive Gaussian Ellipsoid (AGE)	0.408 ± 0.008
Geometric Environment GeoS10	0.150 ± 0.007
Σ vs CC	0.500 ± 0.007
Σ vs AGE	0.509 ± 0.010
CC vs AGE	0.475 ± 0.011
Local Optimal Density (Λ)	0.521 ± 0.007

Table 1. Red Fraction ranges (difference between red fractions at highest and lowest density bins) for different individual environmental measures. The uncertainty estimates are obtained through jackknife resampling. Also shown are ranges when contrasting two different measures of environment. The local optimal density (Λ) has the highest red fraction range.

We first bin the weighted galaxy counts by mass, each bin being of width $\triangle \log(\mathcal{M}/\mathcal{M}_\odot) = 0.1$. Within each mass bin, we assume the counts are bimodally distributed as a function of colour, similar to Baldry et al. (2004a). We therefore search for two peaks (local maxima) within each bin - one each for the red and blue populations. We restrict our search for the blue peak to $0 < u-r < 1.5$ and for the red peak to $1.5 < u-r < 2.25$ - the upper limit of 2.25 is to avoid spurious local maxima that may falsely get identified as real peaks. We then use the midpoint of the two peaks as the separating colour value for that bin. Fitting a line through these midpoints gives us our line of separation for the two populations. The equation of the line of separation used is

$$C = 0.216M - 0.421 \tag{10}$$

where C is the colour $u-r$ and M is the galaxy stellar mass in $\log(\mathcal{M}/\mathcal{M}_\odot)$. Figure 3b shows the results of this approach. The red and blue peaks are identified as dots of the corresponding colour. The red peaks seem to fall on a straight line with limited scatter, and align well with the contour lines. The blue peaks do the same for lower masses $\mathcal{M} \lesssim 10^{10}\mathcal{M}_\odot$. The contours seemed to flatten out in the range $\mathcal{M} \gtrsim 10^{10}\mathcal{M}_\odot, 0 < u-r < 1.5$ and there are hardly any peaks here. We found that varying the different peaks changed the midpoints and the separating line but did not affect the environment signal mentioned above, so we maintain this simple approach. The midpoints and the separating line are shown in black in Figure 3b.

4 RESULTS AND DISCUSSION

In this section, we discuss the results of our investigation. For all plots, we bin the mass and environment variables (except the discrete-valued geometric environment) to increase the signal-to-noise (S/N) ratio of our estimates. Additionally, we omit points (or grids in the heatmaps) that represent fewer than 20 galaxies, to ensure that statistically significant conclusions can be drawn from the results.

4.1 Effect of environment on galaxy colour

Figure 4 shows the variation of galaxy colour with the stellar mass and the environmental densities. As previously mentioned, we use the fraction of galaxies that are red within a given mass-environment bin as our metric for assessing the colour dependence. The general trend from looking at these plots is that both the stellar mass and the environmental density play a key role in explaining the evolution in the red fraction, independent of each other. When looking at a given stellar mass bin, denser regions are dominated by red galaxies. Similarly when looking at regions of a given environmental density, the more massive galaxies are likely to be red.

It is worth asking at this point whether the results in Figure 4 translate from galaxy colour to galaxy star formation rates (SFR). This could be interesting because observational SFR indicators may probe different intrinsic properties of galaxies. We used a specific star formation rate (SSFR) indicator available in the GAMA DR4 `MagPhys` catalog instead of colour. We repeated our analysis using the methodology described in section 3.2 to separate the population into quenched/star-forming subpopulations. We estimated the quenched fraction equivalent to the red fraction and studied its dependence on the environment and mass, similar to Figure 4. Our analysis did not yield any new information compared to the results obtained using galaxy colour. This suggests that the SSFR is not necessarily more reliable at selecting quenched galaxies.

4.2 Comparing different environments

Even though there is some variation in red fraction for all environments in Figure 4, the largest dependence occur for the 5th Nearest Neighbour Surface density Σ and the Cylindrical Count (CC). The effect is weaker for the Adaptive Gaussian Ellipsoid (AGE) density, and the weakest for geometric environments. For a given mass bin, the change in red fraction is less stark as we move from voids to knots, as defined by the Eardley et al. (2015) method.

We can do a quantitative check of this by estimating for each density measure the average red fraction range, i.e. the difference in the red fraction between the lowest and highest density regions. Note that we do not consider the stellar mass variation for this calculation. We can do this because all the environment measures in this approach are computed for the same set of objects and have the same stellar mass range.

The red fraction range is easy to calculate for the geometric measure, which has predefined discrete bins. In contrast, the local environment measures are continuous values, and we must bin the data. To ensure a fair comparison, we choose bins such that the lowest and highest bins contain the same number of objects across the individual environments, and choose the remaining bins to be of the same size of 0.2. For the lowest bin, we set this number to be 25% of all objects in our sample, and for the highest bin, we set it to be 5%. The lowest bin fraction is kept high because the Cylindrical Count measure has a value of 0 - its least possible value - for a quarter of its samples, as mentioned in section 2.5.2. The resulting red fraction ranges are shown in Table 1, with the uncertainty estimates obtained through jackknife resampling. The numbers show that there is varying information across different environmental measures. While the 5th Nearest Neighbour Surface density and the Cylindrical Count density have similar red fraction ranges, the range for the Adaptive Gaussian Ellipsoid density is lower - with $\sim 8\sigma$ difference from the previous two densities. Table 1 also shows the red fraction range for the geometric environment, although we not that it is not directly comparable to the range for local measures since the bin sizes are different.

We can also assess the similarity of information provided by each density measure by looking at how the red fraction compares as a function of two different environmental measurements. Figure

ORIGINAL UNEDITED MANUSCRIPT

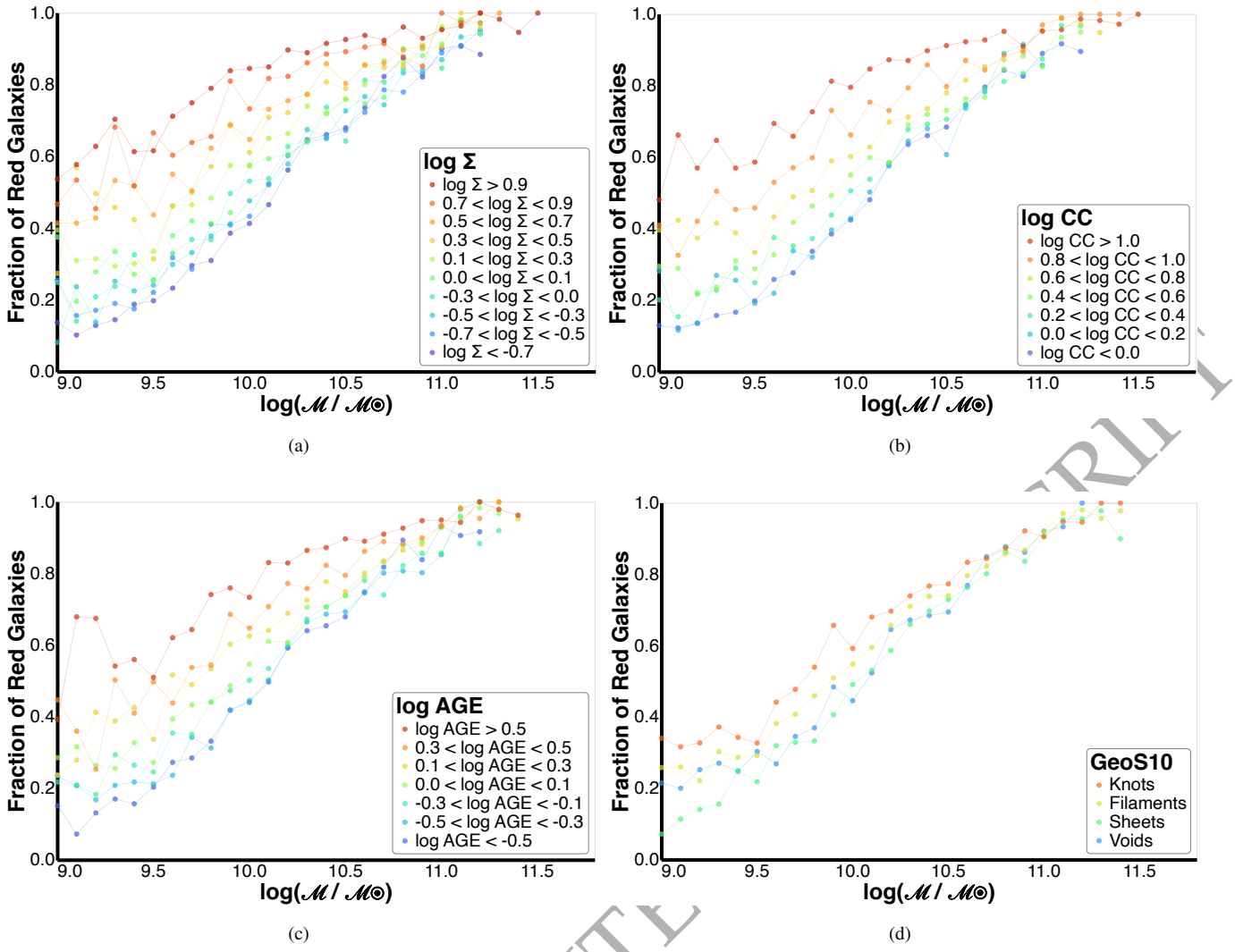


Figure 4. Fraction of red galaxies as a function of stellar mass and different environmental densities. The fractions are computed for bins of stellar mass ($\log(\mathcal{M}/\mathcal{M}_\odot)$, shown on the x-axis) and environment (shown by different colours). The plots are for (a) 5th Nearest Neighbour (Σ), (b) Cylindrical Count (CC), (c) Adaptive Gaussian Ellipsoid (AGE) and (d) Geometric Environment (GeoS10). Note the logarithmic scale for both mass and environmental density.

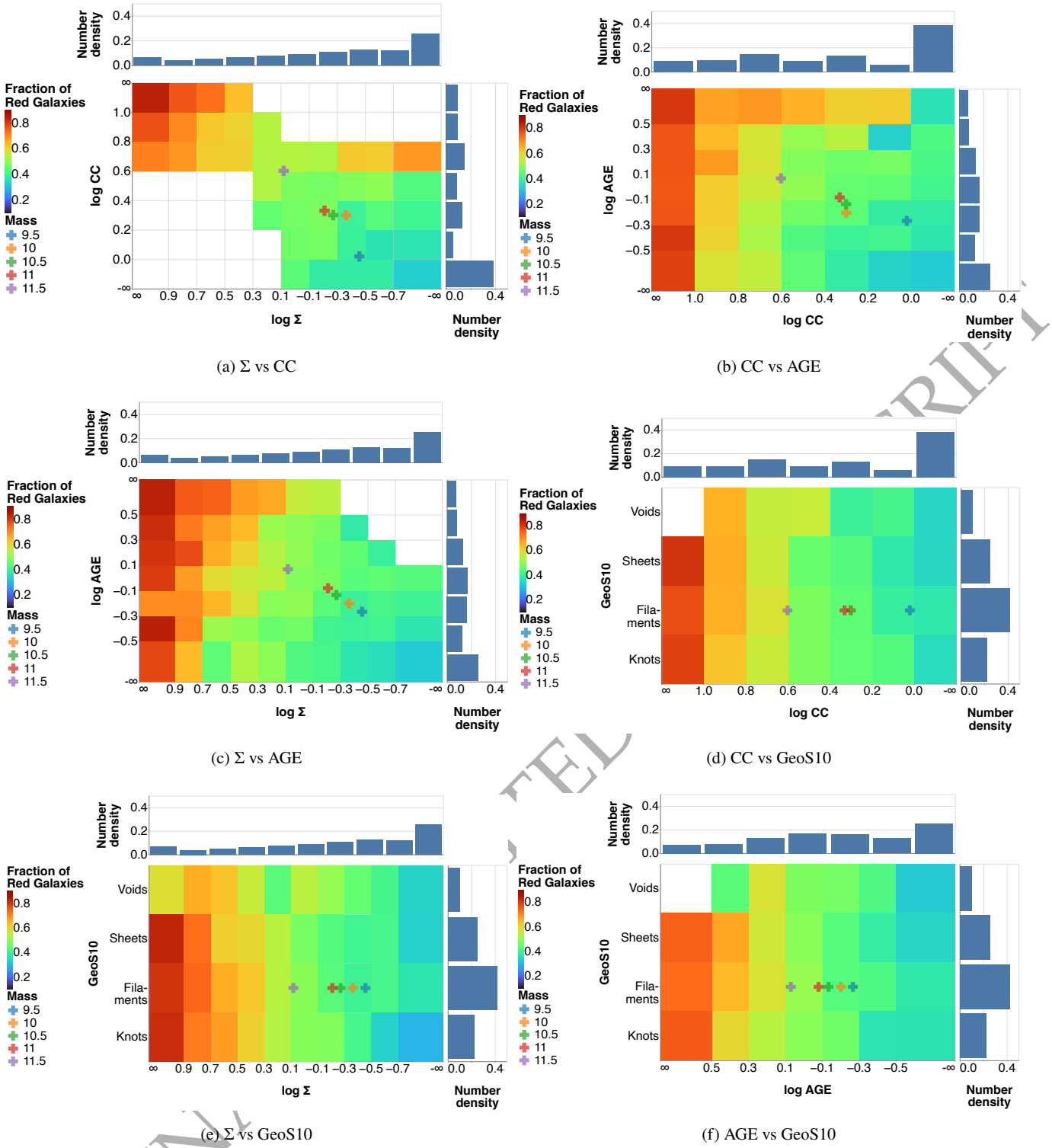
5 shows this as heatmaps. (Here too we are not studying the dependence across stellar mass). Analyzing these plots, we see that there is variation in the red fraction along each environment axis, increasing when moving from lower to higher density regions. Interestingly, if we focus on the heatmaps containing only the local environmental measures (Figures 5a, 5c and 5b), we find that the maximum variation occurs as we move diagonally, from the bottom right (lowest bin for both densities) to the top left (highest bin for both densities). This suggests that all three measures contain different information and are complementing each other. We might therefore be able to define a new environmental measure that combines these three measures and better explains the variation in red fraction. The slightly higher red fraction ranges when comparing the local densities (especially the entries corresponding to Σ vs CC and Σ vs AGE) as shown in Table 1 also support this, although we note that the number of objects in the lowest and highest bins vary by a few percent for these results.

By contrast, if we focus on the heatmaps involving the geometric environment (Figures 5e, 5d, 5f), we see that the variation is lower along the axis of geometric environment. This suggests that the red fraction is relatively less dependent on geometric environment than on local measures.

4.3 Testing the effect of geometric environment

We have seen that the effect on galaxy colour is more pronounced for local environment densities (Figures 4a, 4b and 4c) than the geometric environment (Figure 4d). Here we probe this further by asking whether the geometric environment adds any significant information beyond what is available in the local environment. Testing for this is important for galaxy evolution models. The cosmic web affects galaxy stellar mass (e.g. Kauffmann et al. 2004b) and morphological properties (e.g. Das et al. 2015). It is therefore not unreasonable to expect that large scale flows might influence galaxy evolution. If they do, this information is most likely to be contained in the geometric environment (e.g. Wetzel et al. 2007, 2008; Welker et al. 2020). There is previous work done to identify the effect of large scale environment on dark matter halos (e.g. Hellwing et al. 2021a), suggesting there is interest in the question we seek to answer here.

ORIGINAL UNEDITED



(a) Σ vs CC

(b) CC vs AGE

(c) Σ vs AGE

(d) CC vs GeoS10

(e) Σ vs GeoS10

(f) AGE vs GeoS10

Figure 5. Heatmaps showing the variation in the fraction of red galaxies (represented as colour) as a function of two different environmental densities (one on each axis). Both environmental densities are binned on a logarithmic scale. White cells indicate regions with < 20 samples and where a statistically significant estimate was not achievable. Overlaid on the heatmaps are the median (mode for geometric environment) environmental density values for galaxies in different mass bins. These points show that there is an effect on mass that needs to be controlled for, and we do that in the subsequent analysis (section 4.3). Finally, the sides of each plot show marginalized histograms of individual environmental density, to give information of number density of the galaxies in the different bins.

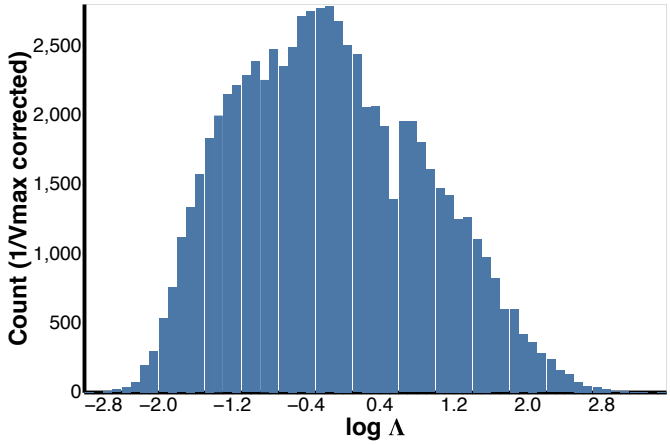


Figure 6. Histogram of local optimal density Λ. We model it as a linear combination of local densities (in log space) that maximizes the red fraction range, thereby best explaining the red fraction variation. Optimizing for this gives $\log\Lambda = \log\Sigma + \alpha\log CC + \beta\log AGE$ with $\alpha = 0.48, \beta = 0.09$. See section 3.1 for meaning of the term $1/V_{max}$ corrected.

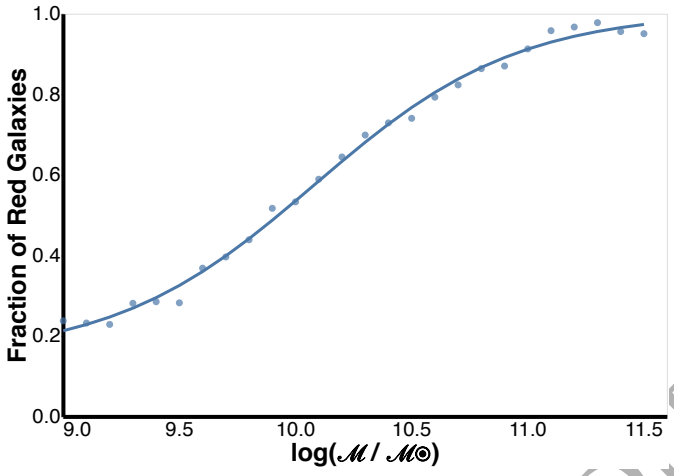


Figure 7. The dependence of the red galaxy fraction on the stellar mass $\log(\mathcal{M}/\mathcal{M}_\odot)$, ignoring (or averaging) the effects of local and geometric environments. The points are the estimations for different mass bins, while the curve is the best fit obtained using the generalized sigmoid function defined in Eqn 12

We define Λ to be the "optimal" local environment measure - the maximal information we can extract out of local environment. Given that the three local environmental measures seem to contain somewhat different information, we model Λ as a linear combination of these (in log space)

$$\log\Lambda = \log\Sigma + \alpha\log CC + \beta\log AGE \tag{11}$$

Note that since our aim is to maximise the red fraction range, the two parameters in equation 11 are sufficient for optimizing - the coefficient on $\log\Sigma$ can be set to 1 without loss of generality.

Quantitatively, optimal density means one which best explains the variation in the red fraction - specifically whose distribution maximizes the red fraction range. We bin the optimal density in the same way we binned the individual densities. To seek the optimal parameters we first did a coarse search of $10^{-3} < \alpha, \beta < 10$, which determined that the optimal parameters lay in the ranges $0.1 < \alpha < 10, 0.01 < \beta < 1$. A subsequent fine grained search of the region yielded best results for $\alpha = 0.48$ and $\beta = 0.09$. For the coarse and fine searches, we found the red fraction range to be continuous in α and β, with only a single notable minima. It makes sense that the optimal coefficient for the cylindrical count is close to 1 (at an order of magnitude level) since it seems to do an equally good job as the nearest neighbour density in explaining the red fraction variation.

Figure 6 shows the histogram of the local optimal density Λ. From Table 1, we see that the maximal red fraction range obtained for Λ is 0.521 ± 0.007. This is 2.2σ higher than the range obtained from the Cylindrical Count density and 2.9σ higher than the range obtained from the 5th Nearest Neighbour Surface density. We therefore recommend using Λ in applications involving precision measurements of the effect of local environment on the red galaxy fraction, e.g. if one aims to remove the local environmental dependence.

In the discussion below, $\tilde{\Lambda}, \tilde{G}, \tilde{\mathcal{M}}$ stand for bins of the local environment ($\log\Lambda$), geometric environment and stellar mass ($\log\frac{\mathcal{M}}{\mathcal{M}_\odot}$) respectively. $r(\boldsymbol{\theta})$ is the red fraction estimate as a function of one or more parameters $\boldsymbol{\theta}$.

We can now look for the effect of geometric environment beyond that accounted for by the local environment and the stellar mass. The environmental density comparisons in Figure 5 show that the median (mode in case of geometric environment) environmental density varies with mass, so we need to control for this effect. We do that by removing the red fraction dependence on stellar mass and examine how the residuals vary with the environment, i.e $r(\tilde{\Lambda}, \tilde{G})$. In order to estimate the stellar mass dependence, we recompute the red fraction values for different stellar mass bins, ignoring the environmental differences , i.e. $r(\tilde{\mathcal{M}})$. The points in Figure 7 show the recomputed red fraction as a function of stellar mass. We found a linear or logistic regression model did not give a good fit to these values, instead we found a generalized sigmoid function to work better (minimized the root mean-squared error)

$$r(\tilde{\mathcal{M}}; a, b, c, d) = \frac{a}{1 + \exp(-b\tilde{\mathcal{M}} + c)} + d \tag{12}$$

The optimal value of parameters obtained was $a = 0.85, b = 2.29, c = 23.15, d = 0.14$. The corresponding curve is shown as the solid line in Figure 7.

To remove the effect of mass from the red fraction estimates, we compute $r(\tilde{\Lambda}, \tilde{G})$ and subtract the average contribution from stellar mass $\langle r(\tilde{\Lambda}, \tilde{G})\rangle_{\tilde{\mathcal{M}}}$. We obtain the latter by averaging $r(\tilde{\mathcal{M}})$ for mass bins of size 0.1, weighted by the number of objects in each $(\tilde{\Lambda}, \tilde{G}, \tilde{\mathcal{M}})$ bin. We label the residuals $R(\tilde{\Lambda}, \tilde{G})$.

$$\begin{aligned} R(\tilde{\Lambda}, \tilde{G}) &\equiv r(\tilde{\Lambda}, \tilde{G}) - \langle r(\tilde{\Lambda}, \tilde{G})\rangle_{\tilde{\mathcal{M}}} \\ &= r(\tilde{\Lambda}, \tilde{G}) - \frac{1}{N}\sum_i r(\tilde{\mathcal{M}}_i) n_i \end{aligned} \tag{13}$$

where n_i is the number of objects in bin $(\tilde{\Lambda}, \tilde{G}, \tilde{\mathcal{M}}_i)$ and $N = \sum_i n_i$.

Figure 8 shows $R(\tilde{\Lambda}, \tilde{G})$ as a function of the local environment (x-axis) and the geometric environment (colour). The error bars shown in the plot are 1-sigma standard error estimates obtained by jackknife resampling. The jackknife resampling is done both when averaging across $r(\tilde{\mathcal{M}})$ and also when computing the red fraction estimates $r(\tilde{\Lambda}, \tilde{G})$. The latter incorporates the uncertainty due to varying sample size across different environment bins. For instance, voids at lowest local densities have ~ 3000 objects, while

ORIGINAL UNEDITED MANUSCRIPT

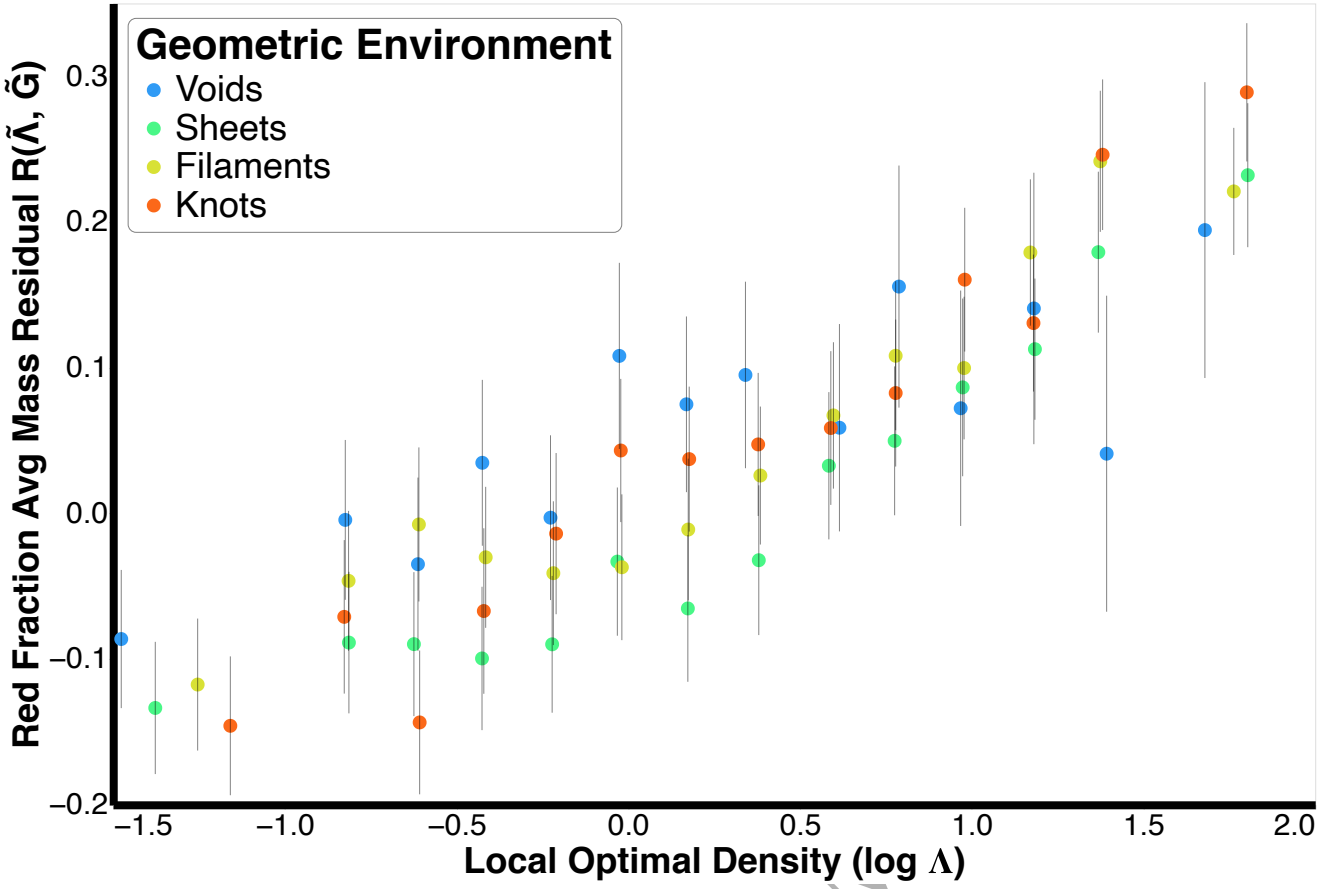


Figure 8. The residual of the red galaxy fraction after subtracting the average stellar mass contribution (based on the fit in Figure 7). The residual is plotted as a function of the local optimal environment log Λ (shown on x-axis) and the geometric environment (shown by various colours). The plot suggests that geometric environment does have an effect on the red fraction. In particular, at low to medium local densities, voids tend to have a larger fraction of their galaxies as red than other geometric environments. The error bars are 1-sigma standard deviation errors obtained using jackknife resampling and incorporate the effect of sample size within each environment bin.

those at the highest densities have ∼ 30 objects. The error bars at higher densities are therefore higher.

The figure suggests a residual geometric environment effect between the different bins, e.g. between voids and sheets. We perform an Analysis of Variance (ANOVA) test to assert whether there is a difference between $R(\tilde{\Lambda}, \tilde{G})$ for the different geometric environment bins when looking at a fixed local environment bin. The results are shown in Table 2 and confirm the presence of a residual difference at significane level $\alpha = 0.1$. The above-average red fraction residuals for voids could be a consequence of the definition of voids as per Eardley et al. (2015) and may require further investigation (see 5.1.1). Conversely, the result may indicate a potential effect of the surrounding large scale structure - perhaps a mechanism for replenishing star-forming gas, whose absence in voids leads to larger fraction of red galaxies. At higher local densities the trend is reversed - voids have lower residuals than other geometric bins, though the error bars are also wider. Finally, the residuals increase as we go from lower to higher local density, reaffirming a positive correlation of red fraction with local environment independent of mass.

5 CONCLUSIONS AND FUTURE WORK

In this work, we studied the effect of environment on the restframe galaxy colour, using the fraction of red galaxies as a metric. We found that the red fraction varies with the local environment independently of the stellar mass. For a given stellar mass, galaxies in a high density environment are more likely to be red when compared to those in a low density environment. Possible explanations include loss of star forming gas due to stripping in denser environments, through mechanisms such as strangulation, galaxy harassment and ram pressure stripping. We also found that the different environmental measures defined in GAMA contain different information. In particular, the local environmental measures exhibit more red fraction variation as we move from regions of lower to higher densities, when compared to the geometric environment. However, when the different local measures are combined to produce an "optimal" local density, there is a residual effect of the geometric environment independent of the local environment and stellar mass.

5.1 Future Work

The following are potential directions for further research:

ORIGINAL UNEDITED

log Λ	Sum of Squares within group	Mean of Squares within group	Sum of Squares between groups	Mean of Squares between groups	F-statistic	df_1	df_2	Significant at $\alpha = 0.1$?
-2.746	38.757	0.002	6.640	2.213	1048.209	3	18354.465	Yes
-0.919	11.689	0.002	3.755	1.251	503.467	3	4701.871	Yes
-0.719	13.116	0.002	12.948	4.316	1584.159	3	4814.153	Yes
-0.519	14.234	0.002	9.468	3.156	1209.381	3	5454.602	Yes
-0.319	14.253	0.002	5.073	1.691	656.601	3	5534.304	Yes
-0.119	13.631	0.002	14.616	4.872	1869.357	3	5229.968	Yes
0.081	11.569	0.002	8.267	2.755	1099.752	3	4616.798	Yes
0.281	9.795	0.002	4.258	1.419	575.594	3	3971.571	Yes
0.481	9.135	0.002	0.535	0.178	65.414	3	3350.898	Yes
0.681	10.407	0.002	2.392	0.797	289.701	3	3780.694	Yes
0.881	8.687	0.002	3.707	1.235	448.259	3	3151.066	Yes
1.081	6.806	0.002	2.355	0.785	313.924	3	2721.003	Yes
1.281	6.459	0.002	9.644	3.214	1186.889	3	2385.002	Yes
1.481	10.863	0.002	7.507	2.502	1145.862	3	4974.112	Yes

Table 2. Analysis of Variance (ANOVA) result to test whether there is a difference in $R(\tilde{\Lambda}, \tilde{G})$ for the different geometric environment bins when looking at a fixed local environment log Λ. The critical F-statistic for $df_1 = 3, df_2 > 120$ at significance level $\alpha = 0.1$ is 3.78. There is therefore a significant residual difference at each local environment bin.

5.1.1 *Investigate the effect of alternate definitions of large scale environment*

The higher value of $R(\tilde{\Lambda}, \tilde{G})$ for voids at a given local density is dependent on the definition of voids within the geometric environment framework of Eardley et al. (2015), i.e. voids being regions surrounded by similar or higher density regions in all directions. It would be useful to repeat our analysis of section 4.3 using alternative means of defining large scale environment, in order to determine whether the apparent strong influence of voids is robust to variations in methodology.

5.1.2 *Studying the residual effect of dark matter haloes on galaxy colour*

The red galaxy fraction is also expected to be affected by dark matter halo mass, with more massive halos expected to have more red galaxies (see e.g. Behroozi et al. 2019). The environmental density certainly plays a role in this relation - more massive dark matter halos would form out of bigger overdensities , which would accrue more baryons, forming more galaxies and thus in general have higher environmental densities. For instance, Baldry et al. (2006) found a correlation between the overdensity and the projected 5th Nearest Neighbour environmental density Σ as estimated from simulations.

It would therefore be interesting to study whether the halo mass has any residual correlation with the red galaxy fraction, after accounting for the effects of environment and stellar mass. If a correlation is found, it may suggest that massive dark matter haloes have some inherent mechanism that affects the evolution of galaxies and makes them more likely to be quenched.

5.1.3 *Studying the environmental effects via simulations*

Future work could also look into seeing how well simulations explain the observed environmental effects. In particular, we would look to see whether simulations also show a small residual dependence on large scale geometric environment, and if so what mechanism is causing this. This would be similar to the approach of Baldry et al. (2006), but for all the environmental densities used in this work. For e.g., the geometric environmental measure is a conceptually different metric, so it would be intriguing to see how well it is reproduced in analytical models, especially given that it has a small effect on the red fraction. Any environmental metric which is measurable in observational data can be reproduced in a cosmological simulation, which is an important step in any robust comparison with simulations. In this case, it would involve identifying mechanisms that explain the observed red fraction trends found in the different environmental measures. It would also involve comparing the different galaxy evolution models with their different feedback mechanisms (e.g. Bower et al. 2006; Somerville et al. 2008; Van Den Bosch et al. 2008; Peng et al. 2015; Bianconi et al. 2016; Bower et al. 2017) to see which models more closely align with the results shown here.

ACKNOWLEDGEMENTS

GAMA is a joint European-Australasian project based around a spectroscopic campaign using the Anglo-Australian Telescope. The GAMA input catalogue is based on data taken from the Sloan Digital Sky Survey and the UKIRT Infrared Deep Sky Survey. Complementary imaging of the GAMA regions is being obtained by a number of independent survey programmes including GALEX MIS, VST KiDS, VISTA VIKING, WISE, Herschel-ATLAS, GMRT and ASKAP providing UV to radio coverage. GAMA is funded by the STFC (UK), the ARC (Australia), the AAO, and the participating institutions. The GAMA website is http://www.gama-survey.org/ .

Our work builds on many software tools/libraries, which we acknowledge by citing. Our analysis was done in the Python ecosystem, with the following packages being the most relevant:

ORIGINAL UNEDITED MANUSCRIPT

- For the environment: Anaconda (Anaconda 2020), Jupyter Notebook (Kluyver et al. 2016)
- For the analysis: Numpy (Harris et al. 2020), Scipy (Virtanen et al. 2020), Pandas (McKinney 2010), Astropy (Astropy Collaboration et al. 2018, 2013), Scikit-learn (Pedregosa et al. 2011), Stan (Riddell et al. 2021; Stan 2021)
- For plotting and visualization: Altair (VanderPlas et al. 2018), Seaborn (Waskom 2021), Matplotlib (Hunter 2007), Plotly (Plotly 2015)

DATA AVAILABILITY

This work uses publicly available data as part of data release 4 (DR4) of the GAMA survey. DR4 can be accessed at http://www.gama-survey.org/dr4/

REFERENCES

Alpaslan M., et al., 2014, Monthly Notices of the Royal Astronomical Society, 438, 177 2
Alpaslan M., et al., 2015, Monthly Notices of the Royal Astronomical Society, 451, 3249 2, 3
Anaconda 2020, Anaconda Software Distribution 14
Arnouts S., et al., 2007, A&A, 476, 137 2
Astropy Collaboration et al., 2013, A&A, 558, A33 14
Astropy Collaboration et al., 2018, AJ, 156, 123 14
Ayromlou M., Nelson D., Yates R. M., Kauffmann G., White S. D. M., 2019, Monthly Notices of the Royal Astronomical Society, 487, 4313 3
Baldry I. K., Glazebrook K., Brinkmann J., Ivezic Z., Lupton R. H., Nichol R. C., Szalay A. S., 2004a, ApJ, 600, 681 2, 6, 8
Baldry I. K., Balogh M. L., Bower R., Glazebrook K., Nichol R. C., 2004b, AIP Conference Proceedings, 743, 106 2, 6
Baldry I. K., Balogh M. L., Bower R. G., Glazebrook K., Nichol R. C., Bamford S. P., Budavari T., 2006, Monthly Notices of the Royal Astronomical Society, 373, 469 2, 3, 5, 13, 15, 16
Baldry I. K., et al., 2012, Monthly Notices of the Royal Astronomical Society, 421, 621 7
Baldry I. K., et al., 2014, Monthly Notices of the Royal Astronomical Society, 441, 2440 3
Baldry I. K., et al., 2018, Monthly Notices of the Royal Astronomical Society, 474, 3875 2, 3, 5
Balogh M. L., Baldry I. K., Nichol R., Miller C., Bower R., Glazebrook K., 2004, ApJ, 615, L101 2
Balogh M. L., et al., 2016, MNRAS, 456, 4364 2
Barsanti S., et al., 2018, ApJ, 857, 71 2
Behroozi P., Wechsler R. H., Hearin A. P., Conroy C., 2019, Monthly Notices of the Royal Astronomical Society, 488, 3143 2, 13
Bell E. F., de Jong R. S., 2001, ApJ, 550, 212 15
Bell E. F., et al., 2004, ApJ, 600, L11 2
Benson A. J., Bower R. G., Frenk C. S., Lacey C. G., Baugh C. M., Cole S., 2003, ApJ, 599, 38 2
Bhowmick A. K., Blecha L., Thomas J., 2020, ApJ, 904, 150 3
Bialas D., Lisker T., Olczak C., Spurzem R., Kotulla R., 2015, A&A, 576, A103 2
Bianconi M., Marleau F. R., Fadda D., 2016, A&A, 588, A105 13
Bond J. R., Kofman L., Pogosyan D., 1996, Nature, 380, 603 3
Bower R. G., Benson A. J., Malbon R., Helly J. C., Frenk C. S., Baugh C. M., Cole S., Lacey C. G., 2006, Monthly Notices of the Royal Astronomical Society, 370, 645 1, 2, 13
Bower R. G., Schaye J., Frenk C. S., Theuns T., Schaller M., Crain R. A., McAlpine S., 2017, MNRAS, 465, 32 1, 13
Brough S., et al., 2013, Monthly Notices of the Royal Astronomical Society, 435, 2903 2, 3, 5
Chabrier G., 2003, PASP, 115, 763 3
Crain R. A., et al., 2015,] 10.1093/mnras/stv725, 450, 1937 1
Croton D. J., et al., 2006, Monthly Notices of the Royal Astronomical Society, 365, 11 1, 2
Das M., Saito T., Iono D., Honey M., Ramya S., 2015, The Astrophysical Journal, 815, 40 9
Davé R., Anglés-Alcázar D., Narayanan D., Li Q., Rafieferantsoa M. H., Appleby S., 2019,] 10.1093/mnras/stz937, 486, 2827 1
Driver S. P., et al., 2009, Astronomy & Geophysics, 50, 5.12 1, 3
Driver S. P., et al., 2011, MNRAS, 413, 971 3
Driver S. P., et al., 2022, Monthly Notices of the Royal Astronomical Society, 513, 439 3, 7
Dubois Y., et al., 2014,] 10.1093/mnras/stu1227, 444, 1453 1
Eardley E., et al., 2015, Monthly Notices of the Royal Astronomical Society, 448, 3665 2, 6, 8, 12, 13
Felten J. E., 1976, The Astrophysical Journal, 207, 700 6
Foltz R., et al., 2018, ApJ, 866, 136 2
Goh T., et al., 2019, MNRAS, 483, 2101 3
Gunn J. E., Gott III J. R., 1972, The Astrophysical Journal, 176, 1 2
Haas M. R., Schaye J., Jeeson-Daniel A., 2012, MNRAS, 419, 2133 3
Hahn O., Porciani C., Carollo C. M., Dekel A., 2007, MNRAS, 375, 489 3
Harris C. R., et al., 2020, Nature, 585, 357 14
Hellwing W. A., Cautun M., van de Weygaert R., Jones B. T., 2021a, Physical Review D, 103, 063517 9
Hellwing W. A., Cautun M., van de Weygaert R., Jones B. T., 2021b, Phys. Rev. D, 103, 063517 3
Hopkins P. F., 2013,] 10.1093/mnras/sts210, 428, 2840 3
Hopkins A. M., et al., 2013, Monthly Notices of the Royal Astronomical Society, 430, 2047 3
Hopkins P. F., Kereš D., Oñorbe J., Faucher-Giguère C.-A., Quataert E., Murray N., Bullock J. S., 2014,] 10.1093/mnras/stu1738, 445, 581 1
Hunter J. D., 2007, Computing in Science Engineering, 9, 90 14
Kauffmann G., White S. D. M., Heckman T. M., Ménard B., Brinchmann J., Charlot S., Tremonti C., Brinkmann J., 2004a, MNRAS, 353, 713 2
Kauffmann G., White S. D. M., Heckman T. M., Ménard B., Brinchmann J., Charlot S., Tremonti C., Brinkmann J., 2004b, Monthly Notices of the Royal Astronomical Society, 353, 713 9
Kluyver T., et al., 2016, in Loizides F., Schmidt B., eds, Positioning and Power in Academic Publishing: Players, Agents and Agendas. pp 87–90 14
Kraljic K., et al., 2018, Monthly Notices of the Royal Astronomical Society, 474, 547 6
Kraljic K., Davé R., Pichon C., 2020, MNRAS, 493, 362 3
Kristensen M. T., Pimbblet K. A., Gibson B. K., Penny S. J., Koudmani S., 2021, ApJ, 922, 127 3
Lagos C. d. P., Bayet E., Baugh C. M., Lacey C. G., Bell T. A., Fanidakis N., Geach J. E., 2012, Monthly Notices of the Royal Astronomical Society, 426, 2142 1
Liske J., et al., 2015, Mon. Not. R. Astron. Soc., 452, 2087 2, 3, 5
Loveday J., et al., 2012, MNRAS, 420, 1239 5
Malavasi N., Langer M., Aghanim N., Galárraga-Espinosa D., Gouin C., 2022, Astronomy and Astrophysics, 658, A113 6
McKinney W., 2010, Proceedings of the 9th Python in Science Conference, pp 56–61 14
Muldrew S. I., et al., 2012, MNRAS, 419, 2670 3
Navarro J. F., Frenk C. S., White S. D. M., 1997, The Astrophysical Journal, 490, 493 1
Pandey B., Sarkar S., 2020, Monthly Notices of the Royal Astronomical Society, 498, 6069 3
Peacock J. A., 2003, AIP Conference Proceedings, 666, 275 1
Pedregosa F., et al., 2011, Journal of Machine Learning Research, 12, 2825 14
Peng Y.-j., et al., 2010,] 10.1088/0004-637X/721/1/193, 721, 193 2, 3
Peng Y., Maiolino R., Cochrane R., 2015, Nature, 521, 192 2, 13
Perez J., Tissera P., Padilla N., Alonso M. S., Lambas D. G., 2009, Monthly Notices of the Royal Astronomical Society, 399, 1157 3
Plotly 2015, Collaborative Data Science, https://plot.ly 14
Reeves A. M. M., et al., 2021, Monthly Notices of the Royal Astronomical Society, 506, 3364 2

ORIGINAL UNEDITED MANUSCRIPT

Riddell A., Hartikainen A., Carter M., 2021, Pystan (3.0.0).Documentation https://pystan.readthedocs.io/, PyPI 14
Robotham A., et al., 2010, PASA, 27, 76 3
Robotham A. S. G., et al., 2011, Monthly Notices of the Royal Astronomical Society, 416, 2640 2, 3
Schaefer A. L., et al., 2017, Monthly Notices of the Royal Astronomical Society, 464, 121 2
Schaefer A. L., et al., 2019, Monthly Notices of the Royal Astronomical Society, 483, 2851 2
Schawinski K., et al., 2007, ApJS, 173, 512 6
Schaye J., et al., 2010,] 10.1111/j.1365-2966.2009.16029.x, 402, 1536 1
Schaye J., et al., 2015,] 10.1093/mnras/stu2058, 446, 521 1
Simpson C. M., Grand R. J. J., Gómez F. A., Marinacci F., Pakmor R., Springel V., Campbell D. J. R., Frenk C. S., 2018,] 10.1093/mnras/sty774, 478, 548 3
Somerville R. S., Hopkins P. F., Cox T. J., Robertson B. E., Hernquist L., 2008, MNRAS, 391, 481 1, 2, 13
Sotillo-Ramos D., et al., 2021, Monthly Notices of the Royal Astronomical Society, 508, 1817 2
Sousbie T., 2011, Monthly Notices of the Royal Astronomical Society, 414, 350 6
Springel V., et al., 2018,] 10.1093/mnras/stx3304, 475, 676 1
Stan 2021, Stan Modeling Language Users Guide and Reference Manual v2.27. https://mc-stan.org 14
Stoughton C., et al., 2002, The Astronomical Journal, 123, 485 2
Strateva I., et al., 2001, The Astronomical Journal, 122, 1861 2, 6
Strauss M. A., 2002, Proceedings of the SPIE, 4836, 1 1
Taylor E. N., et al., 2011b, MNRAS, 418, 1587 5
Taylor E. N., et al., 2011a, Monthly Notices of the Royal Astronomical Society, 418, 1587 2, 5, 15
Taylor E. N., et al., 2015, Monthly Notices of the Royal Astronomical Society, 446, 2144 2, 3, 6, 15
Tecce T. E., Cora S. A., Tissera P. B., Abadi M. G., Lagos C. D. P., 2010,] 10.1111/j.1365-2966.2010.17262.x, 408, 2008 3
Thomas D., Maraston C., Schawinski K., Sarzi M., Silk J., 2010, Monthly Notices of the Royal Astronomical Society, 404, 1775 6
Trayford J. W., Theuns T., Bower R. G., Crain R. A., Lagos C. d. P., Schaller M., Schaye J., 2016,] 10.1093/mnras/stw1230, 460, 3925 3
Trussler J., Maiolino R., Maraston C., Peng Y., Thomas D., Goddard D., Lian J., 2020, Monthly Notices of the Royal Astronomical Society, 491, 5406 2
Tully R. B., Mould J. R., Aaronson M., 1982, The Astrophysical Journal, 257, 527 6
Van Den Bosch F. C., Aquino D., Yang X., Mo H. J., Pasquali A., McIntosh D. H., Weinmann S. M., Kang X., 2008, Monthly Notices of the Royal Astronomical Society, 387, 79 13
VanderPlas J., et al., 2018, Journal of Open Source Software, 3, 1057 14
Virtanen P., et al., 2020, Nat Methods, 17, 261 14
Vogelsberger M., et al., 2014,] 10.1038/nature13316, 509, 177 1
Wang H., Mo H. J., Jing Y. P., Yang X., Wang Y., 2011, MNRAS, 413, 1973 3
Waskom M. L., 2021, Journal of Open Source Software, 6, 3021 14
Welker C., et al., 2020, Monthly Notices of the Royal Astronomical Society, 491, 2864 9
Wetzel A. R., Cohn J. D., White M., Holz D. E., Warren M. S., 2007, The Astrophysical Journal, 656, 139 9
Wetzel A. R., Schulz A. E., Holz D. E., Warren M. S., 2008, The Astrophysical Journal, 683, 1 9
Wiersma R. P. C., Schaye J., Smith B. D., 2009,] 10.1111/j.1365-2966.2008.14191.x, 393, 99 3
Williams R. J., Quadri R. F., Franx M., van Dokkum P., Labbé I., 2009, ApJ, 691, 1879 2
Wilman D. J., et al., 2005, MNRAS, 358, 88 3
Wright A. H., et al., 2016, Monthly Notices of the Royal Astronomical Society, 460, 765 5
Yoon J. H., Schawinski K., Sheen Y.-K., Ree C. H., Yi S. K., 2008, ApJS, 176, 414 2, 6
van de Sande J., et al., 2021, Monthly Notices of the Royal Astronomical Society, 508, 2307 2
van der Burg R. F. J., McGee S., Aussel H., Dahle H., Arnaud M., Pratt G. W., Muzzin A., 2018, A&A, 618, A140 2

APPENDIX A: HOW THE RED FRACTION ESTIMATES VARY AMONG THE DIFFERENT COLOURS

In our analysis above, we used the colour $u-r$ which is obtained from LAMBDAR photometry, k-corrected as well as corrected for Milky-Way dust extinction. We also experimented with two additional colours and studied their dependence on the red galaxy fraction. These two colours are defined below:

- **Based on SDSS Model magnitudes $u_{model} - r_{model}$**: This colour is estimated from model magnitudes provided as part of GAMA from DR7 of the Sloan Digital Sky Survey (SDSS). These are corrected for Milky Way dust extinction. We also perform k-correction to redshift $z = 0$. Figure A1a shows a histogram of the resulting colour values. This colour term is similar to that used by Baldry et al. (2006).
- **Based on GAMA magnitudes with host galaxy dust extinction correction $u^* - r^*$**: The estimates for this colour are obtained similar to $u-r$, but then are additionally corrected for the light absorbed by dust in the host galaxy. This correction is done using stellar population synthesis (SPS) models (Taylor et al. 2011a). In this type of modeling, there can be degeneracy between internal dust reddening and age/metallicity effects. Figure A1b shows a histogram of the resulting colour values. See Taylor et al. (2015) for more details.

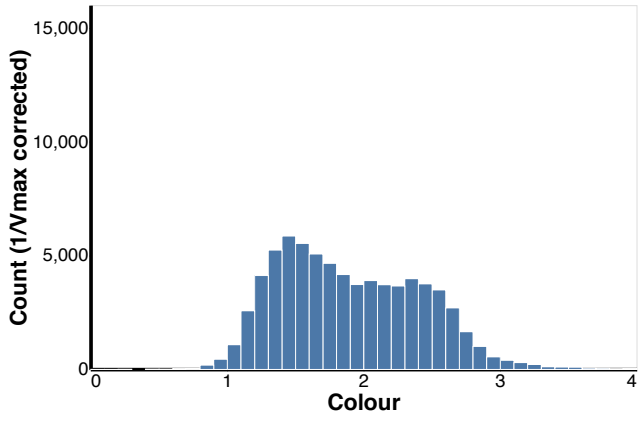
The histogram of $u - r$ is shown again in Figure A1c for comparison. We note that for all the colours, the same approach described in section 3.2 was used to separate the red and blue populations.

Figure A2 shows the resulting red fraction estimates. The plots that use the colours $u-r$ (Figure A2a) and $u_{model} - r_{model}$ (Figure A2b) look similar to each other and to the results obtained by Baldry et al. (2006) (specifically Figure 11(b) of that paper). For high environmental density, we observe high values of red fraction even at low masses. This agreement may be seen as a consistency check for the approach used in this work.

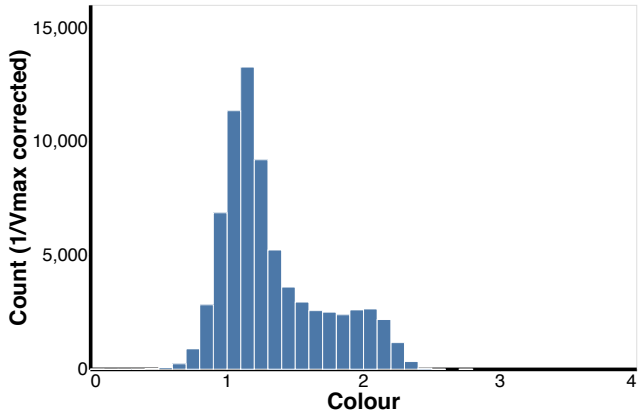
In contrast the plot for the colour $u^* - r^*$ (Figure A2c) has low red fraction values at low masses for high environmental densities. This is different from what has been seen in the literature. As mentioned in 2.3, the colour $u^* - r^*$ is defined similarly to $u-r$, except it also attempts to correct for the extinction caused by dust in the host galaxy (Taylor et al. 2015). This correction involves performing stellar population synthesis (SPS) modeling and is considerably more difficult than estimating the Milky Way dust extinction. Additionally, the degeneracy between dust and the age of a stellar population (Bell & de Jong 2001) makes it difficult to model and extract only the dust part of a galaxy, hence separate models must be fit for the stars and the dust. It is therefore possible that there are some unresolved degeneracies in the SPS fitting and the dust extraction which leads to incorrect values of $u^* - r^*$ colours. This therefore warrants further investigation into how the colour estimates are generated.

The colours $u_{model} - r_{model}$ and $u - r$ produce comparable results, however Taylor et al. (2011a) have shown for SDSS that the magnitudes obtained using model based photometry tend to be biased compared to their Petrosian counterparts, and have therefore recommended using the latter magnitudes. The colour $u - r$ is computed from GAMA magnitudes obtained using LAMBDAR

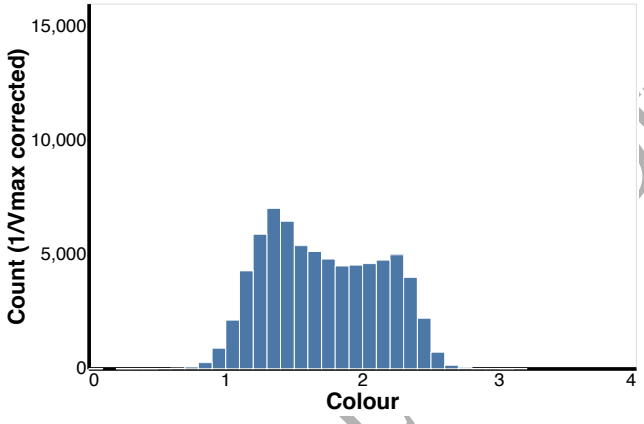
ORIGINAL UNEDITED MANUSCRIPT



(a) $u_{model} - r_{model}$



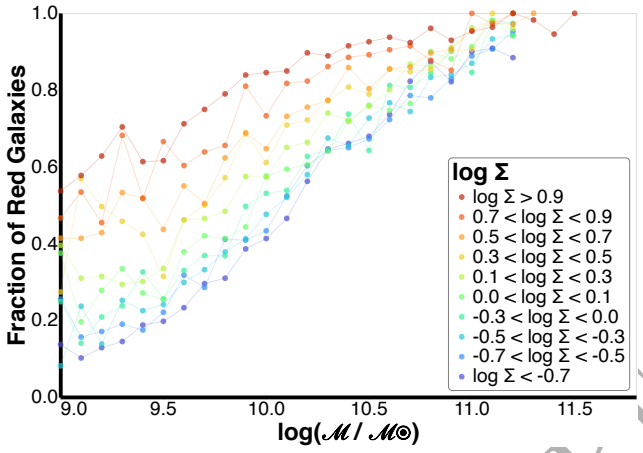
(b) $u^* - r^*$



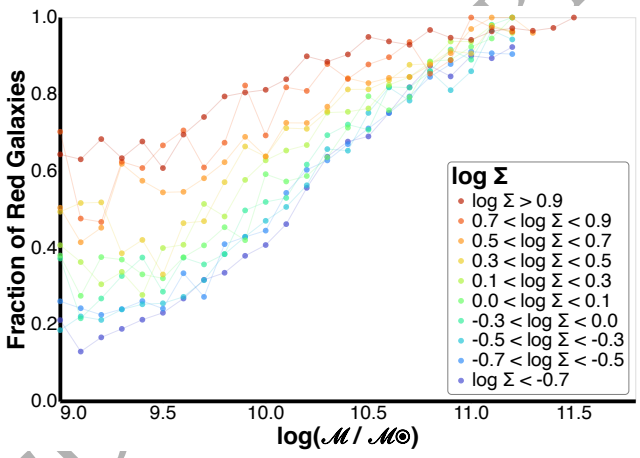
(c) $u - r$

Figure A1. Histograms of the different colours used to study the red fraction dependence - (a) $u_{model} - r_{model}$, (b) $u^* - r^*$ and (c) $u - r$. The counts shown here are volume $(1/V_{max})$ weighted, see section 3.1 for more information on how this correction is done.

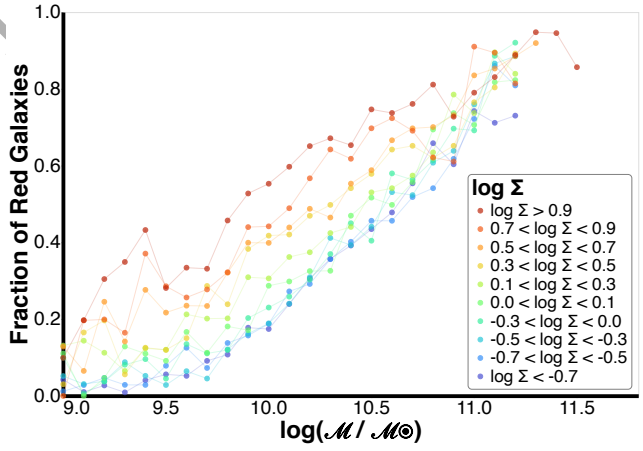
photometry, which uses elliptical apertures rather than circular ones, and is expected to yield even better estimates than Petrosian magnitudes. For this reason, the colour $u - r$ was chosen for all the analyses in this work.



(a) $u - r$



(b) $u_{model} - r_{model}$



(c) $u^* - r^*$

Figure A2. Effect of using different colours on the fraction of red galaxies as a function of stellar mass $\log(\mathcal{M}/\mathcal{M}_\odot)$ and 5th Nearest Neighbour density (Σ). The labels are similar to those shown in Figure 4. Using $u^* - r^*$ results in low red fraction estimates in high environmental regions, especially at low mass. This is contrary to what is known in the literature (Baldry et al. 2006) and warrants further investigation into how the colour estimates are generated.

ORIGINAL ARTICLE